AD_____

Grant Number DAMD17-94-J-4371

TITLE: Computer Aided Breast Cancer Diagnosis

PRINCIPAL INVESTIGATOR: Carey E. Floyd, Ph.D.

CONTRACTING ORGANIZATION: Dale University Medical Center
Durham, North Carolina 27710

REPORT DATE: October 1996

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
October 1996

3. REPORT TYPE AND DATES COVERED
Annual (23 Sep 95 - 22 Sep 96)

4. TITLE AND SUBTITLE
Computer Aided Breast Cancer Diagnosis

5. FUNDING NUMBERS
DAMD17-94-J-4371

6. AUTHOR(S)
Carey E. Floyd, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Duke University Medical Center
Durham, North Carolina 27710

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

The long range goal of this project is to improve the accuracy and consistency of breast cancer diagnosis by developing a Computer Aided Diagnosis (CAD) system for early prediction of breast cancer from the patients' mammographic findings and medical history.

In the second year of this project, we have acquired 250 new cases bring our total case database to over 500. We have investigated two alternative network architectures for predicting malignancy: a genetic algorithm approach, and an adaptive learning rule for the feed-forward network. A user interface was developed for the efficient data entry and error checking was employed. These developments were targeted at the first specific aim of the grant: develop an ANN to predict biopsy outcome from mammographic and history findings. In the third year, we will focus on the second specific aim: evaluate the improvement in radiologists' diagnostic performance when the computer diagnostic aid is provided. This implementation of an accurate CAD system will improve sensitivity, specificity, and consistency of breast cancer diagnosis and will provide a significant improvement in long term outcome for breast cancer patients.

14. SUBJECT TERMS Breast Cancer

15. NUMBER OF PAGES
35

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

(X) Where copyrighted material is quoted, permission has been obtained to use such material.

( ) Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

( ) Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.

( ) In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

( ) For the protection of human subjects, the investigator(s) have adhered to policies of applicable Federal Law 32 CFR 219 and 45 CFR 46.

( ) In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

Carey S. Shields

Principal Investigator's Signature

9 Jan 97

Date

Table of Contents

Progress report on project DAMD17-94-J-4371, "Computer Aided Breast Cancer Diagnosis" for the period 9/23/95-9/22/96.

Introduction

The long range goal of this project is to improve the accuracy and consistency of breast cancer diagnosis by developing a Computer Aided Diagnosis (CAD) system for early prediction of breast cancer from the patient';s mammographic findings and medical history.

While mammography is a sensitive test for early diagnosis of breast cancer, 70% of all the cases which are sent to biopsy are benign. We will develop a CAD system based on Artificial Neural Networks (ANNs) to predict the malignancy of breast lesions from radiologists' reports of the findings from mammograms. The strength of the ANNs for this problem is their ability to learn complex relationships from examples of the data, then to generalize and accurately classify examples which the network has not seen before. This system will learn to predict malignancy by examining a large set of radiographic findings which are paired with biopsy results. The database for this learning will be representative of the patient population. Specifically we will: 1)develop an ANN to predict biopsy outcome from mammographic and history findings and 2) evaluate the improvement in radiologists' diagnostic performance when the computer diagnostic aid is provided. This implementation of an accurate CAD system will improve sensitivity, specificity, and consistency of breast cancer diagnosis and will provide a significant improvement in long term outcome for breast cancer patients.

What follows is a point by point assessment of the progress for each task in the original statement of work:

Statement of Work

Task 1, Develop an ANN to predict biopsy outcome from mammographic and history findings.

Years 1-4

Development will start with the successful preliminary backpropagation network. The significant improvements needed include: 1)larger set of clinical cases to better represent the general patient population, 2)higher specificity while maintaining >98% sensitivity. The preliminary work will be extended as follows.

Year 1

1.1)Expand the number of input features, both mammographic and medical history. The ANN will be implemented on a workstation (SUN SPARC) to allow the size of the network to be enlarged. This will allow more medical history and radiological features to be included.

These tasks were all achieved in year one.

Year 2-4

1.2)Develop a time-series ANN to examine current as well as previous exams.

Not yet achieved.

1.3)Evaluate other ANN architectures which have been demonstrated to be appropriate for pattern classification.

Partially achieved in year 2. Documented below.

Year 3-4

2)Evaluate the improvement in radiologists' diagnostic performance when the computer diagnostic aid is provided.

Partially achieved in year one.

Year 3

Install the trained network on the Mammography Database server to perform on-line prediction as the radiologists input the features.

Partially achieved in year 2. Interface built.

Year 3-4

Test the hypothesis that use of the network prediction by radiologists will increase diagnostic accuracy (prediction of biopsy results).

Not yet achieved. To be begun in year 3.

In summary, we have achieved all work for year one, some of the work allocated to years 2-4, some of the work allocated to year 3, and some of the work allocated to year 3-4. We feel that we are ahead of schedule and that we will complete all work by the end of year 4.

In the second year of the grant we have published two peer-reviewed manuscripts[1][2]] with three more accepted for publication[3][4][5]. There have been 3 presentations with published proceedings at professional meetings [7][8][9]. Specifically, we have 1)acquired 240 new cases using the standardized BI-RADS reporting system, 2)

All of this work has been specifically directed toward the first specific aim of the proposal.

In summary:

	Year 2:	Cumulative
Peer-reviewed manuscripts published or in press:	4	6
Published Conference Proceedings:	2	6
International Meeting presentations:	3	10
Related grants received:	1	2

Peer-reviewed manuscripts published or in press:

1 Baker JA, Kornguth PJ, Lo JY, Floyd CE Jr.: An Artificial Neural Network Approach to Improve the Quality of Breast Biopsy Recommendations *Radiology*;**198**;131-135; 1996.

2 Baker JA, Kornguth PK, Floyd CE Jr.: Bi-rads Standardized Mammography Lexicon: Observer Variability of Lesion Description *Amer. J. Roent.* , Apr 1996.

3 Tourassi GD, Floyd CE Jr."The Effect of Data Sampling on the Performance Evaluation of Artificial Neural Networks in Medical Diagnosis". (Accepted in *Medical Decision Making* 1996).

Floyd CE Jr., Lo JY, Tourassi GD, Baker JA, Vitittoe NF, Vargas-Voracek R: Computer Aided Diagnosis in Thoracic and Mammographic Radiology. Medical Imaging Technology, 6;629-634;1996.

Published Conference Proceedings:

Floyd CE Jr, Tourassi GD, Baker JA: Use of genetic algorithms for computer-aided diagnosis of breast cancer from image features. In Proceedings of the International Society for Optical Engineering (SPIE) 2710:51-58, (1996).

2 Lo JY, Kim J, Baker JA, and Floyd CE, Jr, "Computer-aided diagnosis of mammography using an artificial neural network: Predicting the invasiveness of breast cancers from image features," Medical Imaging 1996: Image Processing, Loew MH, Ed., SPIE Medical Imaging 1996: Image Processing, Proc. SPIE 2710: 725-732 (1996).

Meeting presentations (in addition to those listed in the conference proceedings):

2. Lo JY, Baker JA, Kornguth PJ, Floyd CE Jr. .Computer-aided diagnosis of mammography:Artificial neural networks for optimized merging of standardized BIRADS features. Presented at *World Congress on Neural Networks,* International Neural Network Society Annual Meeting (INNS), 1996.

Narrative:

In the second year of the project, we continued the development of an artificial neural network (ANN) to assist radiologists in the differentiation of benign from malignant lesions. Inputs to the ANN were derived from the patient's history and the radiologist's description of lesion morphology following the ACR Breast Imaging Reporting and Data System (BI-RADS^{tm}). The output of the neural network is the likelihood of malignancy.

Artificial neural networks are a form of artificial intelligence analogous to layers of biological neurons. These networks can be trained to "learn" essential information from a set of data. The structure of an ANN is a set of processing units (nodes) arranged in rows. Input nodes are interconnected by simple calculations with an internal layer of hidden nodes and a single output node . Rather than having a fixed algorithmic approach to a classification problem, an ANN is sequentially presented with a set of supervised training cases — input data paired with the correct output. The ANN modifies its behavior ("trains") by adjusting the strength or "weights" of the connections until its own output converges to the known correct output. The information "learned" by the ANN is stored in the weight the network gives to connections between nodes.

ORGANIZATION OF THE NEURAL NETWORK

The ANN for prediction of breast malignancy was constructed as a three layer feed-forward network with a backpropagation training algorithm. The layers consist of an input layer with 18 input nodes, one hidden layer with 10 nodes, and an output layer with one output node. Each input node corresponds to either a radiologist's description of a feature of the lesion or information from the patient's medical or family history.

Of the 402 women undergoing needle localization for nonpalpable breast lesions between January 1991 and December 1992, 194 mammograms were randomly selected from a list of patient history numbers for prospective evaluation. A total of 206 lesions were identified on these studies that went on to open excisional biopsy and pathological diagnosis.

Each set of mammograms was acquired using film-screen technique on dedicated mammography equipment. No case was included in the study if either of the reviewing radiologists had prior knowledge of the biopsy results or if the suspicious area was not definitely identified. Of the 206 lesions evaluated there were 99 masses alone, 76 suspicious calcifications, and 11 combinations of masses and associated microcalcifications. The remaining 20 lesions included various combinations of architectural distortion, regions of asymmetric breast density, areas of focal asymmetric density, and areas of asymmetric breast tissue. Patients ranged in age from 24 to 86 years with an average age of 55 years. At biopsy, 133 (65%) of the lesions were found to be benign while 73 (35%) were malignant. This PPV of 35% is somewhat greater than that described in prior studies. Twenty-four of the 73 malignancies (33%) were not yet invasive.

Each set of training films was reviewed prospectively by one of two radiologists whose primary clinical responsibilities are the interpretation of mammograms and the evaluation of breast lesions and who are familiar with the definitions of the BI-RADS^{tm} descriptors. One radiologist evaluated 151 cases while the other reviewed the remaining 55 cases (206 total cases). At least two views of the breast with the

suspicious lesion were provided to the participating radiologists; a cranio-caudal and mediolateral-oblique view were available in all cases. Other views including true lateral, magnification views, and spot compression views as well as comparisons with the opposite breast were provided for evaluation when available. In order to avoid biasing the radiologist's description of the lesion, films from prior studies and the patient's history were initially withheld while the reviewing radiologist chose descriptors for each lesion. The radiologist was asked to describe each lesion using the BI-RADS^{tm} lexicon by completing a checklist that included all possible BI-RADS^{tm} descriptors. The reviewing radiologist was permitted to select only a single descriptor from each category. Each reader was blinded to the biopsy results while reviewing the films. The lesion descriptors along with patient history were used as inputs to train a neural network as described below.

Finally, to compare the performance of the ANN to experienced radiologists, the reviewing mammographer was provided with the patient's history and any prior films to correlate with the study mammograms and was requested to estimate the likelihood of malignancy. A five point scale was used with 1=very likely benign, 2=likely benign, 3=indeterminate, 4=likely malignant, and 5=very likely malignant.

NETWORK INPUTS

A total of 18 inputs were used to train the ANN to distinguish benign from malignant lesions. Ten of the inputs consisted of morphologic features extracted from the lesion by a radiologist. The remaining 8 inputs encompassed data from the patient's personal and family history collected from a survey form completed by the patient at the time of the exam. Each input is information routinely collected using the ACR BI-RADS^{tm} standardized lexicon.

The first three features are descriptive features that apply to microcalcifications and calcifications associated with masses: calcification distribution, number and description. Inputs four through seven apply only to masses: mass margin, mass shape, mass density, and mass size. Three descriptive features that can apply to all

lesions include lesion location, associated findings (e.g. axillary adenopathy), and special cases (e.g. asymmetric breast tissue).

The remaining 8 inputs are data from each patient's history. These include the patient's age, history of prior breast cancer, history of prior ipsilateral benign biopsy, weak, intermediate or strong family history of breast cancer, menstrual status, and use of estrogen or progesterone therapy. Each morphologic feature and patient history data was assigned a numerical value which was then scaled so that each input ranged from zero to one. The order of the inputs in each category was determined at the beginning of the study by discussion with experienced mammographers and review of reports discussing the malignant potential of various BI-RADS^{tm} descriptors.

Prediction of Mammography Biopsy Outcome Using Neural Networks: alternate backpropagation architectures

Introduction

The data used for this project was assembled previously. Previous of our publications (described in the year one report) documented the performance of a two-layer neural-network in predicting the outcome of biopsy. This project was to evaluate the performance of different neural network architectures. The previous work used a two layer backpropagation network and after using a cross-validation style training strategy, ROC areas in the range of 0.89 were achieved on the dataset. The objective for this project was to meet or exceed the performance of the single hidden node network.

Background

The data for 205 mammography patients (previously collected) was used. The results of these surveys were recorded, along with some history data for each patient. The history information included: menopausal status, age, prior biopsy results for previous visits, hormonal supplements, and family history. A total of 18 findings were used to train the network.

Methods

This project was prototyped, debugged and executed in the Matlab mathematical programming environment. Matlab supplies this thing called the neural network toolbox, but it's training methods did not have the capability of observing error on an arbitrary test set while training. Therefore, matlab's backpropagation training software was modified to plot and record the error on a test set while training. Also incorporated was the ability to display ROC curves and their areas as training progresses to get some measure of how the network is behaving during training. ROC area and the shape of the ROC curve are very important indices for this study. The ROC area calculator portion of the code was calibrated and tested using the LabRoc software and the matlab *.m* file *genlabroc.m* and it proved to be accurate to within 2 percent which was deemed close enough.

We theorized that adding a third hidden layer to the network would greatly increase generalization ability. It was hoped that the network would have better performance on input data that it had never been trained on. Since there were 18 inputs and a relatively large number of training examples, a 3 layer network could be capable of more complex mappings and therefore would be more capable of generalizing than the 2 layer network used previously. It was desired to maximize the network's ability to generalize, so a training strategy documented in the literature called cross-validation was used.

In cross validation, the data set is evenly split into n-different bins then trained and tested several different times on combinations of examples in different bins. So, for example if a data set consists of 20 examples and we'd like to use 5-fold cross validation, we'd split the dataset into 5 bins of 4 examples each. Then for the first iteration of training, we'd test on the first 4 examples, and train on the last 16 and

record the epoch number where we saw that the error on the test set was a minimum. During the second iteration, we'd train on the first four and the last twelve, leaving four in the middle for testing. We'd keep doing it until we'd trained and tested on every combination of "bins". This training technique yields a good approximation of how many epochs to train the network so it will generalize the best when trained on all of the examples (no test-set).

Results

The 2 layer network obtained performance ROC areas in the range of 0.89. The 3 layer networkprovided ROC areas around 0.92. Several different networks were trained using cross-validation techniques and some of the results are summarized in the following table.

Run	N-fold c.v.	Input Neurons	Neurons in H.L. 1	Neurons in H.L. 2	Output Neurons	Epochs trained	ROC Area
A	5	18	10	3	1	577	0.9246
B	5	18	10	3	1	584	0.9235
C	3	18	10	3	1	545	0.9232
D	3	18	3	3	1	784	0.9227
E	3	18	30	3	1	461	0.9233

Discussion

Runs A and B were done to check the sensitivity of the final ROC area on the random number generators used to generate the initial conditions for the neural network weights. The effects seem to be negligible. The effects of using 3-fold cross validation instead of 5 fold cross validation were explored by comparing runs B and C. The time savings from using 3-fold versus 5-fold were considerable, so the minimal difference in ROC areas for the same networks was a welcome result. The remainder of the simulations for this project were done using 3 fold c.v. instead of 5-fold. Runs D and E explored using different numbers of hidden nodes in the neural network to determine it's sensitivity to changes in architecture. Again, the 3 layer network outperformed the 2 layer net by several percent and was shown to be relatively insensitive to changing the number of neurons.

The results of this project are very encouraging; We will continue to work on this problem into the third year, exploring different neural networks.

Genetic algorithm

PURPOSE

In this investigation we have explored genetic algorithms as a technique to train the weights in a feed-forward neural network to predict breast cancer from mammographic findings and patient age.

METHODS

Mammograms were obtained from 206 patients who obtained breast biopsies. Mammographic findings were recorded by radiologists for each patient. In addition, the outcome of the biopsy was recorded. Of the 206 cases, 73 were malignant while 133 were benign at the time of biopsy. A genetic algorithm (GA) was developed to train a feed-forward artificial neural network (ANN) so that the ANN would predict the outcome of the biopsy when the mammographic findings were given as inputs. The GA is a technique for function optimization that mimics biological genetic evolution. The ANN was a fully connected feed-forward ANN with 11 inputs, one hidden layer with 10 nodes, and one output node (benign/malignant). The GA approach allows much flexibility in selecting the cost function to be optimized. In this work three functions were explored as optimization criteria: 1)mean-squared error (MSE); 2) area (Az) under the receiver operating characteristic (ROC) curve; and 3)specificity at a fixed sensitivity of 95%. The system was trained using round-robin sampling.

RESULTS

Optimizing for MSE and Az result in different solutions. The "best" solution was obtained by minimizing a linear combination of MSE and (1-Az). ROC areas 0.82 +/- 0.05 were not significantly different from those obtained using backpropagation for ANN training (0.90 +/- 0.05).

CONCLUSIONS

A new technique for computer-aided diagnosis of breast cancer has been explored. The flexibility of the GA approach allows optimization of cost functions that have relevance to breast cancer prediction.

INTRODUCTION

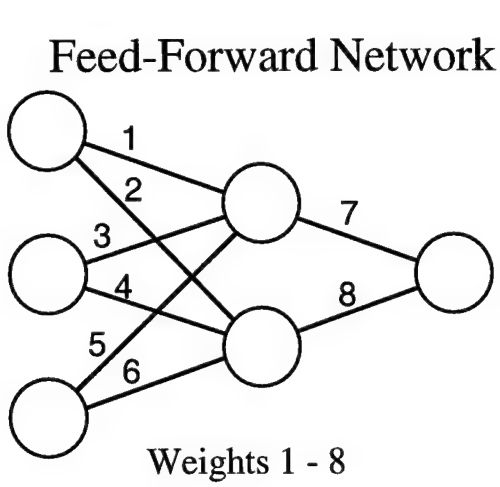
In this investigation we have explored genetic algorithms as a technique to train the weights in a feed forward neural network designed to predict breast cancer based on mammographic findings and patient history. The novel advantage of this technique is the ability to optimize the system for maximizing ROC area rather than minimizing mean squared error.

METHODS

Mammograms were obtained from 206 patients who obtained breast biopsies. Mammographic findings were recorded by radiologists for each patient. In addition, the outcome of the biopsy was recorded. Of the 206 cases, 73 were malignant while 133 were benign at the time of biopsy. Details of this data set have been previously described. A genetic algorithm (GA) was developed to adjust the weights of an artificial neural network (ANN) so that the ANN would output the outcome of the biopsy when the mammographic findings were given as inputs.

The GA is a technique for function optimization that reflects biological genetic evolution.

Here, the GA is implemented to find the weights of a feed forward artificial neural network. Following the biological analog, the individual weights of the ANN are considered to be genes in a strand of DNA. This correspondence is shown below for a network with 8 weights: thre inputs, two hidden nodes, and one output node.



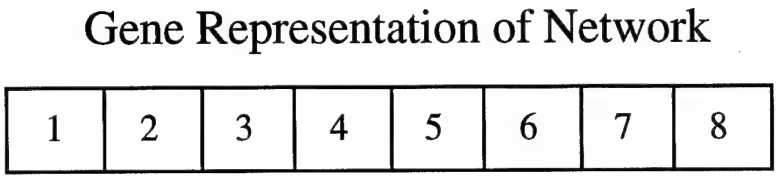


Fig. 1 The correspondence between the feed forward network weights and the genetic representation of the network.

The GA approach is to produce improved networks by using genetic operations on a pool of candidate networks represented as strands of DNA. The genetic operations are: reproduction, cross-linking, and mutation. A fitness criteria is defined and reproduction allows the fittest to survive. The rational for the GA approach is that some weights in each candidate network have a high fitness, and some do not. cross-linking allows survivors from the previous generation to exchange genetic material. Mutation allows new weight values to be introduced into the genetic pool. The iterations of the algorithm are called generations to follow the biological analog.

There are seven user-specified parameters of this model: 1)the number of hidden nodes in the network, 2)the number of candidate networks that make up the breeding pool, 3)the number of networks that survive into the next generation, 4)the cross-over rate that specifies what fraction of the potential number of cross-over sites are used, 5)the mutation rate that specifies what fraction of the total number of weights are randomly modified at each generation, 6)the mutation range that specifies

the maximum percentage of the variation to be applied to those weights that are selected for mutation, and 7)the total number of generations.

For our example, the breeding pool consists of six networks. All networks in the pool have the same architecture and the same number of hidden nodes. With this restriction, the networks may be represented as lists of weights. A single iteration of the evolution of this system is described schematically below in fig. 2

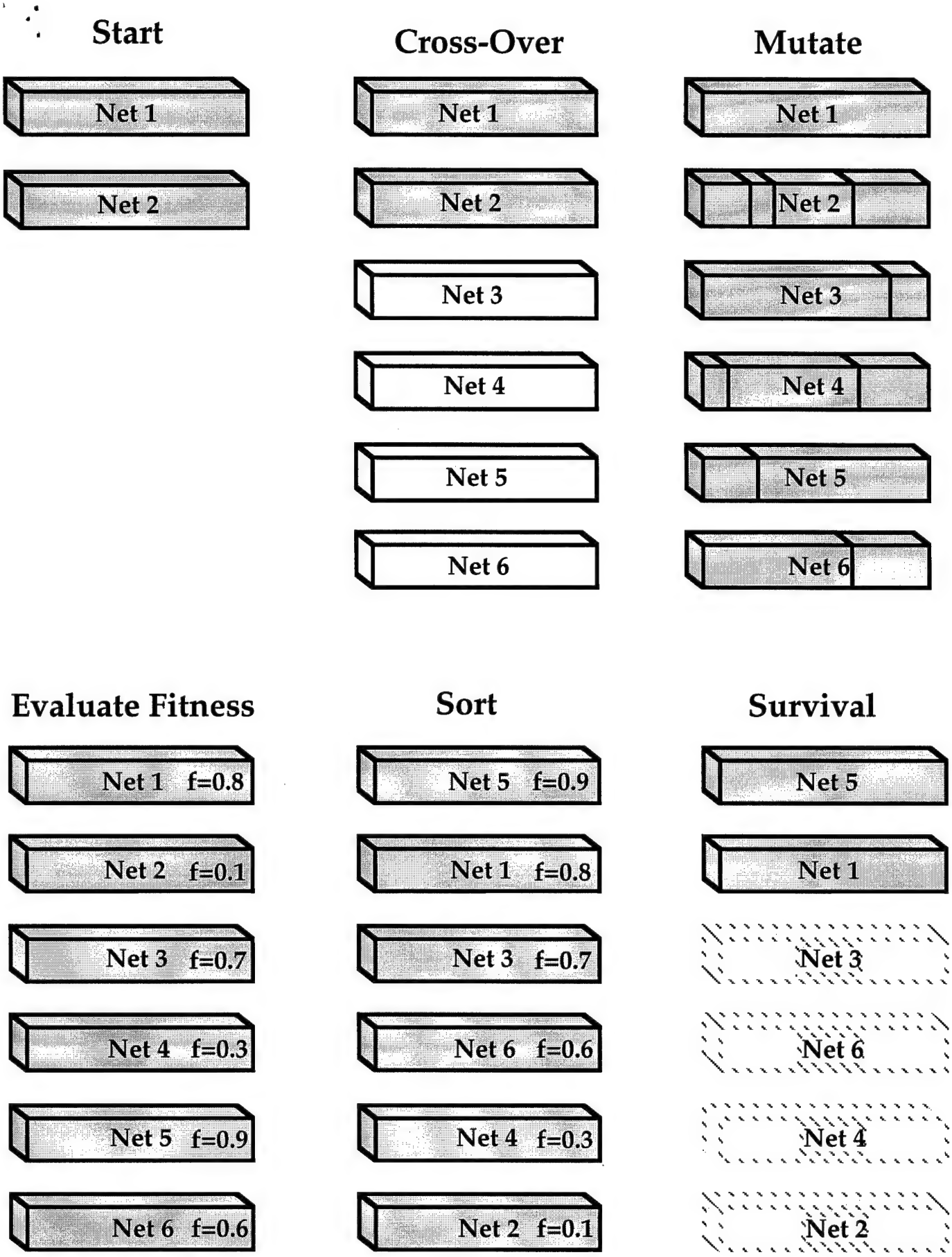


Fig. 2 Schematic of one generation in the evolution of the genetic algorithm.

Two networks are allowed to survive at each generation. The algorithm begins by initializing the weights for all starting networks to random floating point values between -0.3 and 0.3. This selection of limits for the starting values is arbitrary but effective. The cross-over operation is then performed on these networks (represented as lists of weights) to form the remaining 4 members of the pool. Mutation is applied to randomly selected weights throughout the pool. The most fit member of each generation is not mutated to ensure the survival of the best one from each generation. This ensures that the algorithm will converge: the fitness function is improved or at worst stays the same as generations progress. This technique does allow the risk that a solution with an initially good fitness will dominate the evolution, but including other survivors does reduce this risk. Each of the 6 networks is then evaluated for fitness. They are sorted in decreasing order of fitness and the top 2 are selected to become the starting members of the next generation. In this manner, each generation starts with the best from the previous generation and forms new combinations of weights from those that survived. These new combinations are formed through cross-linking (described below). Cross-linking only can swap weights among the members of the breeding pool. New values are introduced through the mutation operation. This iterative process continues until stopped.

Cross-linking is achieved by randomly selecting two of the surviving networks denoted network A and network B. A random location is selected then two new networks are formed by combining the weights up to the location from network A(B) and the remaining weights from network B(A). This opperation is shown for the example network with 8 weights in fig. 2.

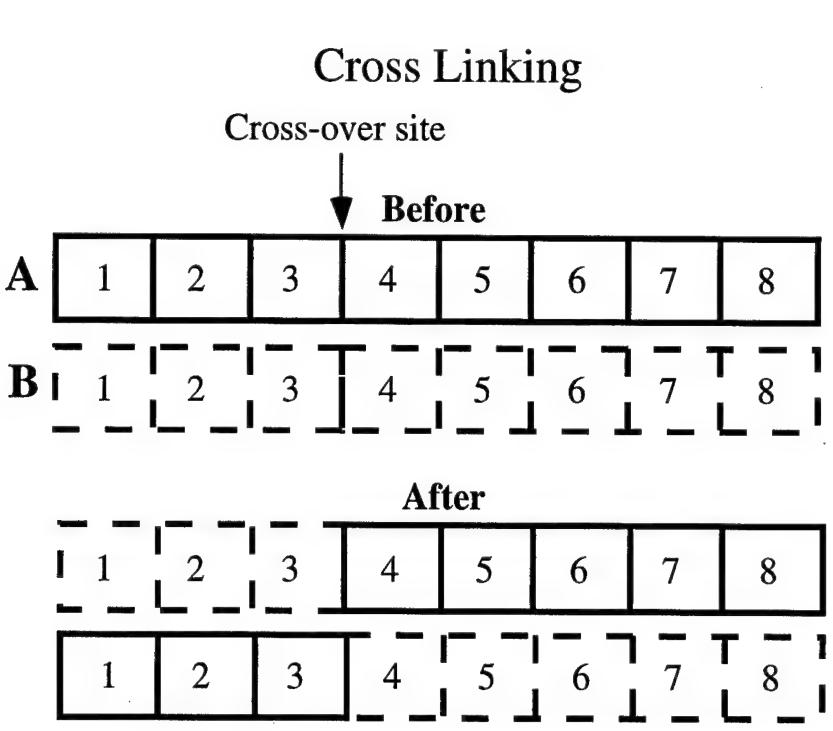


Fig. 2 Example of the cross-linking operation for a network with eight weights. Here the cross-over site is selected after the third weight.

In this example, the cross-over site is selected after the third weight.

Mutation is achieved by randomly selecting a weight to be mutated and then modifying its value by a random percentage. The range over which the percentage is randomly generated is specified as an input.

The GA approach allows much flexibility in selecting the function to be optimized. In this work both mean-squared error (MSE) and receiver operating characteristic (ROC) curve area (Az) were explored as optimization criteria. Fitness is defined as a linear combination of MSE and ROC area. While in principal the fitness function is to be maximized, here the minimum operation is used. Since a minimization criteria is used, we actually evaluate $(1 - A_Z)$. The fitness function is defined as

$$\text{fitness} = a\,(\text{mse}) + (1-a)(1-A_Z)$$

eq. 1

where a is a user-specified weighting constant. For a = 1, the fitness function is the mse. For a = 0, the fitness function is $(1-A_Z)$. Thus for a = 0, minimizing the function will maximize Az.

The ANN for the breast cancer problem is a fully connected feed-forward network using a sigmoid activation with 11 inputs, one hidden layer with 10 nodes, and one output node (benign/malignant). When trained using backpropagation, the performance of this network structure has been described in detail previously. The inputs are either a radiologist's description of a feature of the mammogram or information from the patient's medical history. Here we have selected 10 features from the mammogram plus the patient's age.

CASE SELECTION

A total of 194 mammograms were randomly selected from a list of 402 women who had needle localization for nonpalpable breast lesions between January 1991 and December 1992. From these, 206 lesions were identified that went on to open excisional biopsy and pathological diagnosis. This pathological diagnosis was taken as the gold standard.

All mammograms were acquired using film-screen technique with a grid. No case was included in the study if either of the reviewing radiologists had prior knowledge of the biopsy results or if the suspicious area was not definitely identified. Of the 206 lesions, 99 had masses alone, 76 had calcifications alone, and 11 had combinations of masses and microcalcifications. The remaining 20 cases had combinations of architectural distortion, asymmetric breast density, focal asymmetric density, and asymmetric breast tissue. The age of the patients ranged from 24 to 86 years with an average age of 55 years. As determined by biopsy, 133 (65%) of the lesions were benign while 73 (35%) were malignant.

Radiologists read the mammograms for the selected cases and reported their observed findings using a standard reporting lexicon: BI-RADStm. The input categories are listed below in table 1.

Table 1 - Input Features into the Neural Network

BIRADS Lesion Descriptors			**BIRADS Lesion Descriptors**		
Input Node	Feature	Finding	Input Node	Feature	Finding
1	Calcification Distribution		8	Location	
		no calcifications			
		diffuse			axillary tail
		regional			posterior
		segmental			middle
		linear			anterior
		clustered			subareolar
					central
2	Calcification Number				
		no calcifications	9	Associated Findings	
		< 5			none
		5 to 10			skin lesion
		> 10			hematoma
					trabecular thickening
3	Calcification Description				nipple retraction
		no calcifications			skin retraction
		milk of calcium-like			skin thickening
		rim			architectural distortion
		skin			axillary adenopathy

		vascular
		spherical
		suture
		coarse
		large rod-like
		round
		dystrophic
		punctate
		indistinct
		pleomorphic
		fine branching
4	Mass Margin	
		no mass
		well
		circumscribed
		microlobulated
		obscured
		ill-defined
		spiculated
5	Mass Shape	
		no mass
		round
		oval
		lobulated
		irregular

Input Node	Special Cases
10	none
	intramammary lymph node
	asymmetric breast tissue
	focal asymmetric density
	tubular density or
	solitary dilated duct

Features from medical History

Input Node	Feature
11	Age

6 Mass

Density

no mass

fat-containing

low density

isodense

high density

7 Mass Size

The trained system was evaluated using a round-robin sampling procedure. One of the 206 examples was removed from the original set. The GA was trained on the remaining 205 cases and tested on the removed case. The removed cases was then replaced and another case was removed. This process was repeated until each case had been selected for testing. The test results for each individual case were combined. This sampling proceedure has two advantages for estimating the general performance of an ANN on cases that it has not seen in training: 1)first, for each run there is no overlap between training and testing data; 2)second, since the ANN is trained on N-1 cases, it is very close to the network that would result from training on all N cases. This second point is important since a network is only representative of the data set on which it is trained.

RESULTS

Since an exaustive grid search of 7 parameters using a round-robin of 206 cases is not currently feasible, we chose to fix some of the operating parameters of the system based on empirical experience. After preliminary calculations (and from

previous experience with backpropagation training on this data set), we chose to fix the number of hidden nodes at 10. After several hundred experiments with user guidence, we found stable performance (defined as: little change in results with small variations of the parameters) with the following parameters.

Number of inputs		11
Number of hidden nodes		10
Number of networks in breeding pool		30
Number of survivors at each generation	8	
Probability of cross-linking for a given site		0.03
Probability of mutation for a given weight		0.05
If mutated, maximum range of mutation		100%
Number of iterations	15	
Total number of weights per network		131

Of particular interest was the relative weighting of mse and Az in the fitness function. The bootstrap average mse and Az are plotted below in fig. 3 for 7 values of this weighting. The "fraction of mse" is the value of the coefficient a in the fitness function described by equation 1 above. Note the dramatic improvement in Az as the Az fraction is increased from 0 to 10%.

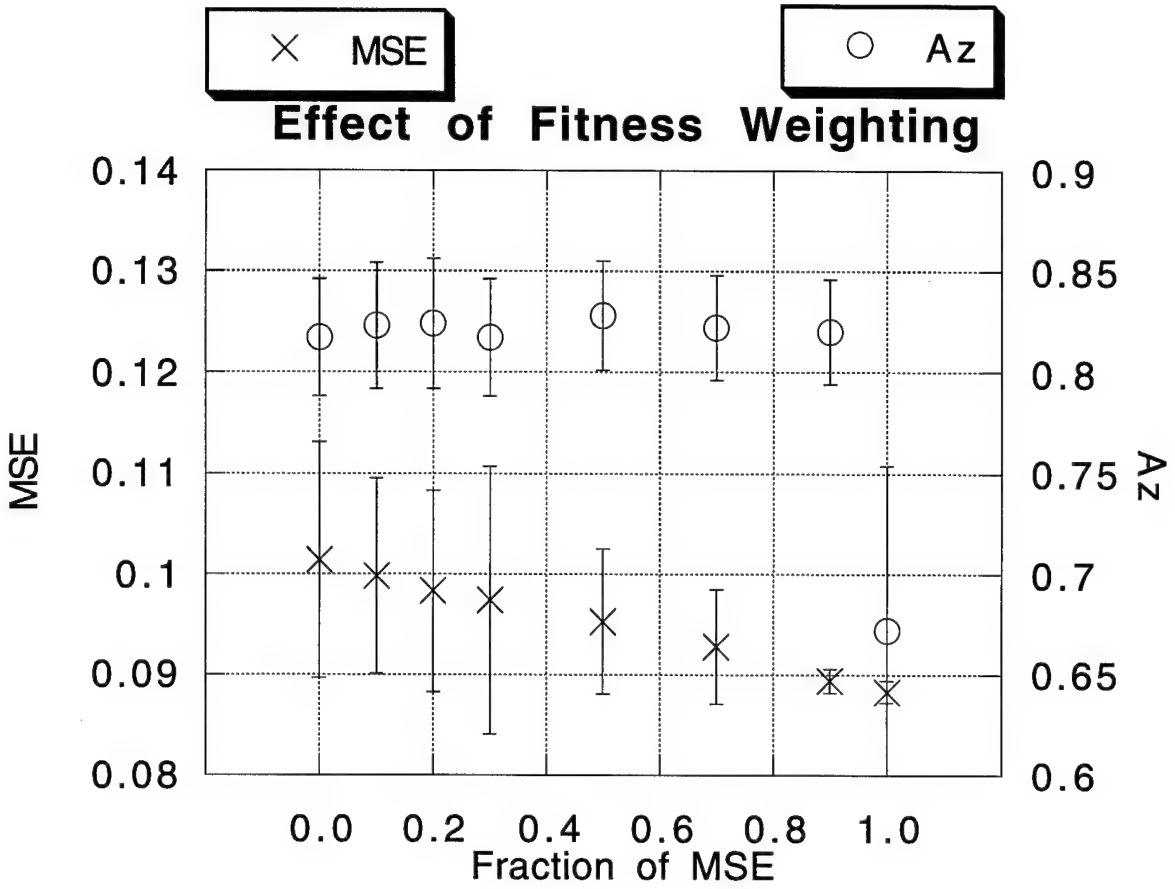


Fig. 3. Performance of the trained network as the fitness function is changed from Az only (Fraction of mse = 0) to all mse (Fraction of mse = 1.0).

The histogram output for the GA optimizing 90%mse and 10% Az is shown in fig. 4.

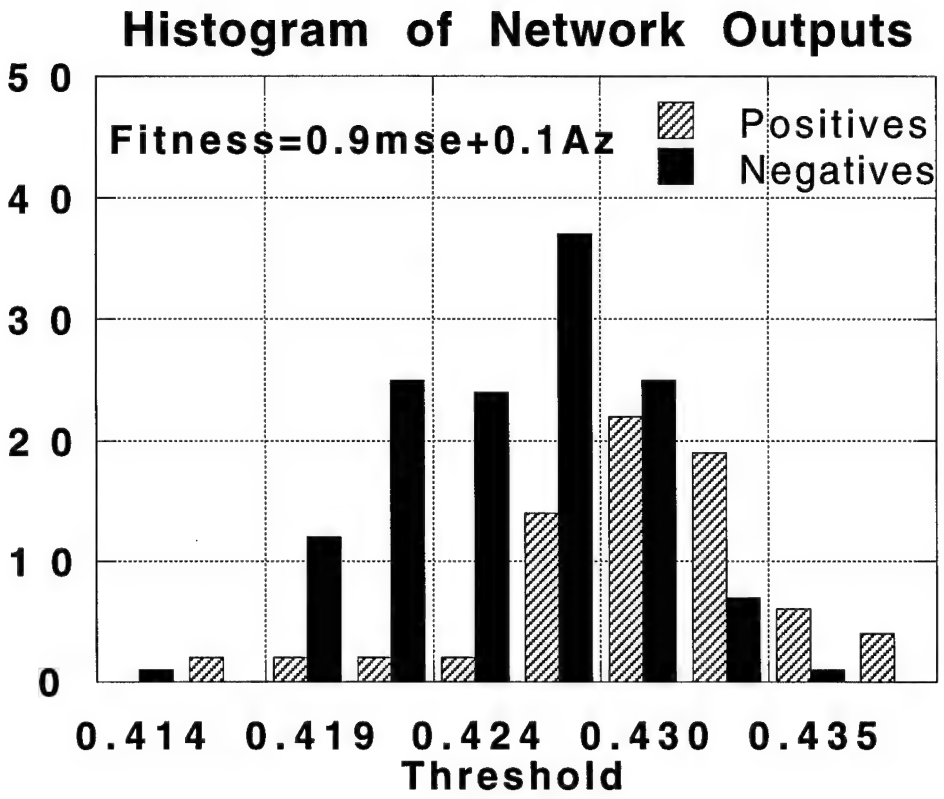
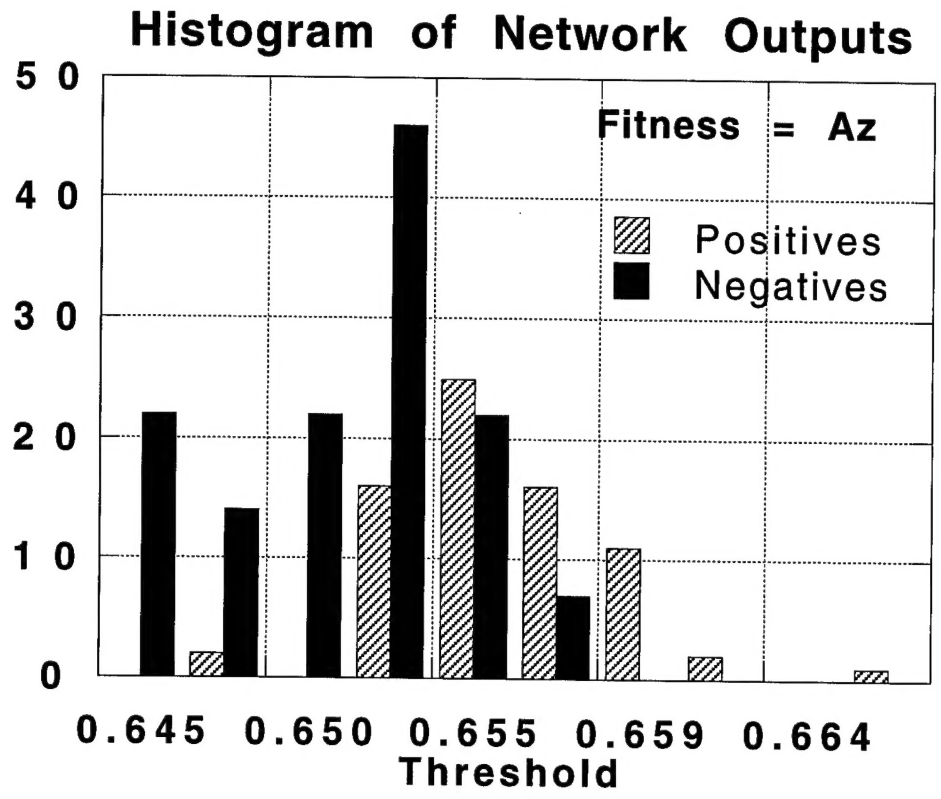
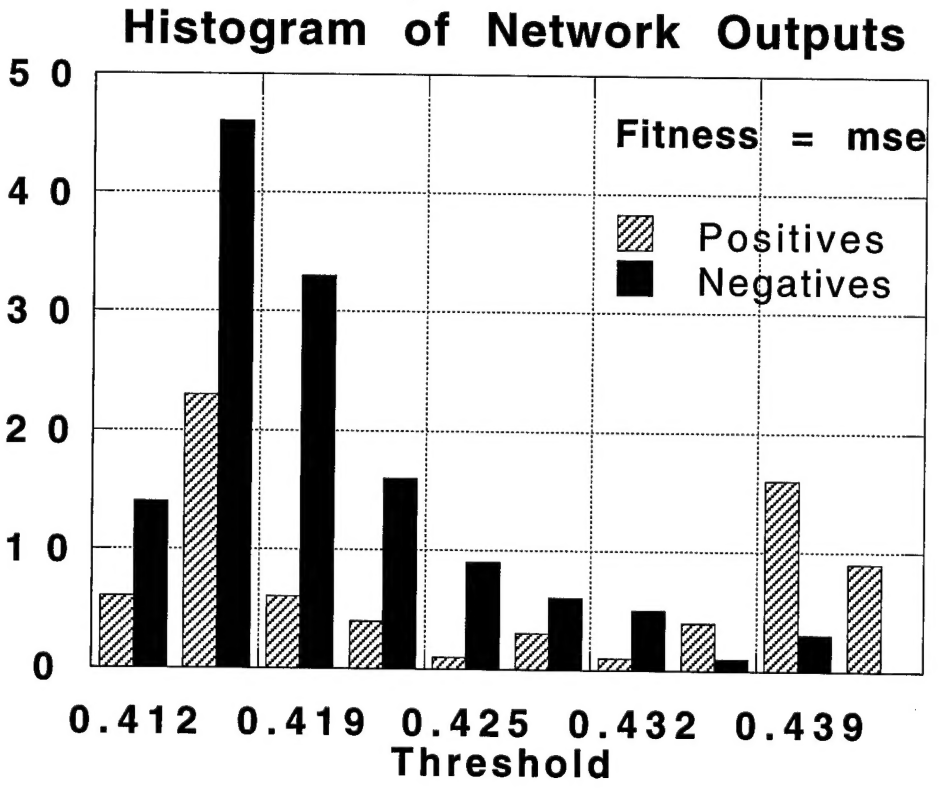


Fig. 4. The histogram of network outputs for the malignant (positive striped bars) and benign (negative solid bars) cases. The fitness function was 90% mse and 10% Az.


Histogram of Network Outputs
50
40
30
20
10
0
Fitness = Az
Positives
Negatives
0.645 0.650 0.655 0.659 0.664
Threshold


Histogram of Network Outputs
50
40
30
20
10
0
Fitness = mse
Positives
Negatives
0.412 0.419 0.425 0.432 0.439
Threshold

For comparison, the histogram output for backpropagation training of the same network structure is shown in fig. 5. The difference is obvious. even though both solutions have about the same Az, The actual histograms are distributed over quite a different range of output.

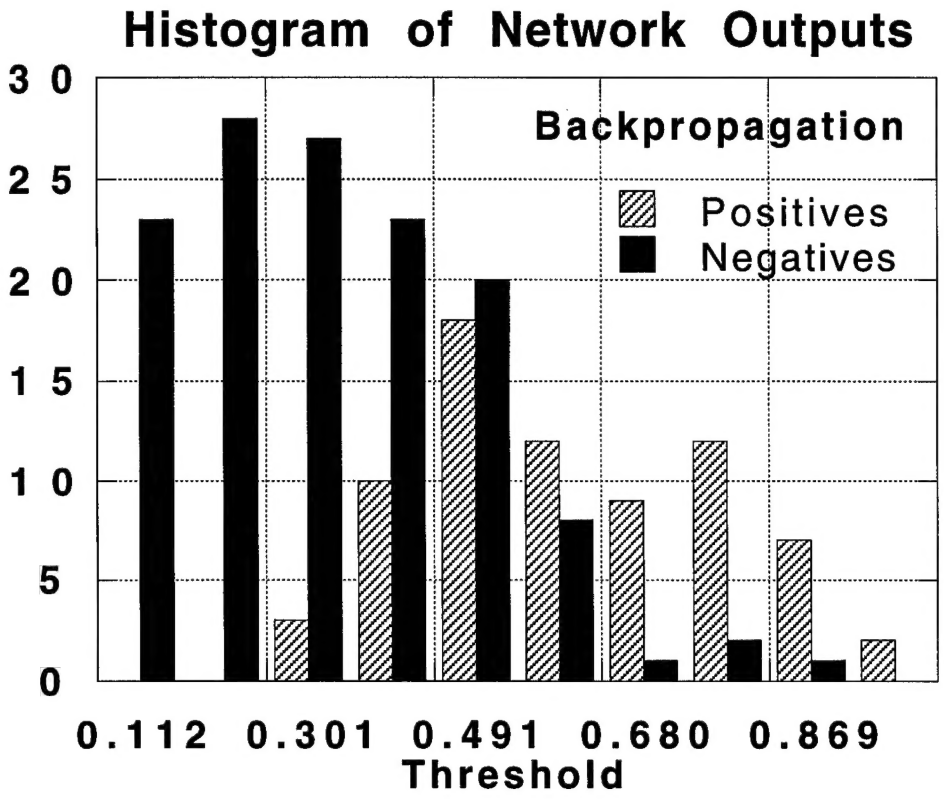


Fig. 5. The histogram of network outputs for the malignant (positive striped bars) and benign (negative solid bars) cases. Here the network was trained by backpropagation which minimizes mse alone.

Receiver Operating Curves (ROC) were calculated using a non-parametric Newtonian integration. The ROC areas calculated for the bootstrap averages were

Fitness	Az	mse
mse	0.67 +/- 0.08	0.088 +/- 0.001
0.9 mse + 0.1 Az	0.82 +/- 0.03	0.089 +/- 0.001
Az	0.82 +/- 0.03	0.101 +/- 0.012
Backpropagation	0.90 +/- 0.05	0.050 +/- 0.033

CONCLUSIONS

A new technique for computer-aided diagnosis of breast cancer has been explored. The flexibility of the GA approach allows optimization of cost functions that have relevance to breast cancer prediction. The performance of the GA trained system was sightly poorer than the same network structure when trained with the gradient-decent backpropagation technique. The result that the mse for the GA was not as low as for backpropagation (even when the GA fitness was exclusively mse) suggests that the GA had not reached an optimum.

Budget

While the research associate has not been hired, other qualified researchers have been partially funded to carry out this work until an appropriate research associate is hired.

Difficulties

No new difficulties have been identified. One difficulty was described last year. The original statement of work was based on the use of a computer database of mammographic findings. Since the time of submission, the clinical use of this database has changed. In addition, as we developed our data acquisition protocol, we found that some items that we needed were not available from the database. While we are negotiating the modification of the on-line data entry forms, we have been acquiring data using paper forms. These forms do not constitute much additional work for the mammographers and have been received with acceptance. We have acquired BI-RADS findings for every biopsy case for the last year. This paper-based data collection system is in place and we anticipate no interruption of data acquisition for the duration of the grant. Since the study section identified the on-line database as a strength of the grant proposal, we continue to actively work to straighten out the compromises required to achieve the on-line data collection. In truth, the difference between paper-based and on-line data collection has no effect on the scientific quality of the research. To rectify this difficulty, we applied for and were awarded a supplement to this grant by the National Action Plan on Breast Cancer to develop and install a computerized data acquisition system. The interface for this system has been developed (in Year 2) and is in final testing prior to being installed in the clinical work area.